

CLAIMS

1. A method of preparing a relational database having a many-to-one relationship for data mining, the method comprising the steps:

5 generate a hierarchical data tree based on a relational data model and

 perform a bottom-up summarization starting from the children and proceeding to the next higher level.

2. A method of including many records in a child level
10 with one record in a parent level for data mining, the method comprising the steps:

 identify a parent level record;

 select child-level records corresponding to the parent level record;

15 characterize the child-level records into a transformed field; and

 append the transformed field to the parent-level record.

3. The method according to claim according to claim 2 wherein the transformed field is one of a plurality of
20 transformed fields.

4. The method according to claim according to claim 2 further comprising the steps:

 provide a record class;

25 provide a characterizing function associated with the record class; and

 categorize the selected child as members of the record class;

 wherein the step categorize step uses the characterizing function to determine the transformed field.

30 5. The method according to claim 3 wherein provide a record class step includes the steps:

provide as a first class time series records with a regular sampling interval, the characterizing function associated with the first class of records being a selected from the group of digital signal processing algorithms
 5 consisting of local cosine transform coefficients and linear predictive coding coefficients; ;

provide as a second class time series records having an irregular sampling interval, the characterizing function associated with the second class of records begin selected
 10 from the group consisting of trend analysis, Markov modeling, and statistical summarization;;; and

provide as a third class of miscellaneous records having no apparent time dependence, the characterizing function associated with the third class of records being selected from
 15 the group consisting of statistical summarization and data association.

6. A method of preparing a relational database for data-mining as a flat database, the method comprising the steps:

generate a hierarchical data tree based on a relational
 20 data model;

perform a bottom-up summarization of the data scattered across multiple tables; and

use a single table containing the summarized data for data mining.

25 7. A method of preparing a relational database for data-mining as a flat database, the method comprising:

identify a data model;

generate a data hierarchy tree;

collect multiple events in child records associated with
 30 a parent record;

characterize the nature of multiple events in the child record;

extract features from the child records, where feature extraction depends on the nature of the multiple events in the child records;

- append extracted features to the parent record; and
5 repeat the method for all child records.

8. A method for transforming a relational database to a flat database, the method comprising the steps:

- provide a relational database having a first table and a second table, each table having a plurality of records, each
10 record having a plurality of fields, wherein a linked field in a selection record in the first table contains data corresponding to data in a linking field of a plurality of records in the second table;

- characterize the data in a summarized field in the second
15 table by computing summarization data, wherein the summarized field in the second table is not the linking field; and

append a summarization field to the first table; and
store the summarization data in the summarization field of the selection record in the first table.

- 20 9. The method according to claim 8 further comprising the step: repeat the characterizing step and the appending step for all records in the first table.

10. A method of applying a data mining technique for a flat database to a relational database, the method comprising
25 the steps:

provide a relational database having a parent table, parent-table records, a child table, and child-table records, wherein a plurality of child table records can be linked to a parent table record;

- 30 convert the relational database to a flat database by appending to a parent table record at least one field

summarizing the values in child table records linked to the parent table; and

apply a flat database data mining technique to the flat database.

- 5 11. A method to determine the relationships among tables in a database, the method comprising the steps:

identify potential primary key fields;

determine table hierarchy that identifies tables as parent tables and related child tables;

- 10 explore intra-table data relationships to reduce the size of a data table; and

explore inter-table data relationships between data in a parent table and data in a child table to that parent.

- 15 12. A method to identify potential primary key fields, the method comprising the steps:

identify a redundant field whose name appears in a plurality of tables;

- 20 identify as a parent table a table in which the value of the redundant field is unique for each record, whereby the redundant field is a primary key for the parent table;

select as a parent record a record from the parent table, whereby the value of the redundant field of the parent record is unique in the parent table;

- 25 select as child records all records in tables other than the parent table for which the value of the redundant field is the same as the value of the redundant field in the parent record; and

identify as a child table a table that is not the parent table and that has the redundant field.

- 30 13. A computer system that can prepare a relational database having a many-to-one relationship for data mining, the computer system comprising:

a means for generating a hierarchical data tree based on a relational data model and

a means for performing a bottom-up summarization starting from the children and proceeding to the next higher level.

5 14. A computer system that can include many records in a child level with one record in a parent level for data mining, comprising:

a means for identifying a parent level record;

a means for selecting child-level records corresponding
10 to the parent level record;

a means for characterizing the child-level records into a transformed field; and

a means for appending the transformed field to the parent-level record.

15 15. The computer system according to claim according to claim 14 further comprising:

a means for providing a record class;

a means for providing a characterizing function associated with the record class; and

20 a means for categorizing the selected child as members of the record class;

wherein the means for categorizing uses the characterizing function to determine the transformed field.

25 16. The computer system according to claim 15 wherein the means for providing a record class further comprises:

a means for providing as a first class time series records with a regular sampling interval, the characterizing function associated with the first class of records being a selected from the group of digital signal processing
30 algorithms consisting of local cosine transform coefficients and linear predictive coding coefficients;

a means for providing as a second class of time series records having an irregular sampling interval, the characterizing function associated with the second class of records begin selected from the group consisting of trend
5 analysis, Markov modeling, and statistical summarization., and

a means for providing as a third class of miscellaneous records having no apparent time dependence, the characterizing function associated with the third class of records being selected from the group consisting of statistical
10 summarization and data association.

17. A computer system that can prepare a relational database for data-mining as a flat database, comprising:

a means for identifying a data model;
a means for generating a data hierarchy tree;
15 a means for collecting multiple events in child records associated with a parent record;

a means for characterizing the nature of multiple events in the child record;

a means for extracting features from the child records,
20 where feature extraction depends on the nature of the multiple events in the child records;

a means for appending extracted features to the parent record; and

a means for repeating the method for all child records.

25 18. A computer system that can transform a relational database to a flat database, comprising:

a means for providing a relational database having a first table and a second table, each table having a plurality of records, each record having a plurality of fields, wherein
30 a linked field in a selection record in the first table contains data corresponding to data in a linking field of a plurality of records in the second table;

a means for characterizing the data in a summarized field in the second database by computing summarization data, wherein the summarized field in the second database is not the linking field; and

- 5 a means for appending a summarization field to the first table; and

a means for storing the summarization data in the summarization field of the selection record in the first table.

- 10 19. . A computer system that can apply a data mining technique for a flat database to a relational database, the comprising:

a means for providing a relational database having a parent table, parent-table records, a child table, and child-
15 table records, wherein a plurality of child table records can be linked to a parent table record;

a means for converting the relational database to a flat database by appending to a parent table record at least one field summarizing the values in child table records linked to
20 the parent table; and

a means for applying a flat database data mining technique to the flat database.

20. A computer system that can determine the relationships among tables in a database, the method
25 comprising the steps:

a means for identifying potential primary key fields;

a means for determining table hierarchy that identifies tables as parent tables and related child tables;

- a means for exploring intra-table data relationships to
30 reduce the size of a data table; and

a means for exploring inter-table data relationships between data in a parent table and data in a child table to that parent.

21. A computer system that can identify potential
5 primary key fields, comprising:

a means for identifying a redundant field whose name appears in a plurality of tables;

a means for identifying as a parent table a table in which the value of the redundant field is unique for each
10 record. whereby the redundant field is a primary key for the parent table;

a means for selecting as a parent record a record from the parent table, whereby the value of the redundant field of the parent record is unique in the parent table;

15 a means for selecting as child records all records in tables other than the parent table for which the value of the redundant field is the same as the value of the redundant field in the parent record; and

a means for identifying as a child table a table that is
20 not the parent table and that has the redundant field.

22. A computer readable medium article of manufacture with instructions for the purpose of preparing a relational database having a many-to-one relationship for data mining, the medium comprising instructions that when executed:

25 generate a hierarchical data tree based on a relational data model and

perform a bottom-up summarization starting from the children and proceeding to the next higher level.

23. A computer readable medium article of manufacture
30 with instructions for the purpose of including many records in a child level with one record in a parent level for data mining, the medium comprising instructions that when executed:

identify a parent level record;
select child-level records corresponding to the parent
level record;
characterize the child-level records into a transformed
5 field; and
append the transformed field to the parent-level record.

24. The computer readable medium according to claim 23
wherein the transformed field is one of a plurality of
transformed fields.

10 25. The computer readable medium according to claim 23,
further comprising instructions that when executed:

provide a record class;
provide a characterizing function associated with the
record class; and

15 categorize the selected child as members of the record
class;

wherein the step categorize step uses the characterizing
function to determine the transformed field.

26. The computer readable medium according to claim 25,
20 wherein the instructions that when executed provide a record
class further comprises instructions that when executed:

provide as a first class time series records with a
regular sampling interval, the characterizing function
associated with the first class of records being a selected
25 from the group of digital signal processing algorithms
consisting of local cosine transform coefficients and linear
predictive coding coefficients;

provide as a second class of time series records having
an irregular sampling interval, the characterizing function
30 associated with the second class of records begin selected
from the group consisting of trend analysis, Markov modeling,
and statistical summarization, and

provide as a third class of miscellaneous records having no apparent time dependence, the characterizing function associated with the third class of records being selected from the group consisting of statistical summarization and data association.

27. A computer readable medium article of manufacture with instructions for the purpose of preparing a relational database for data-mining as a flat database, the medium comprising instructions that when executed:

generate a hierarchical data tree based on a relational data model;

perform a bottom-up summarization of the data scattered across multiple tables; and

use a single table containing the summarized data for data mining.

28. A computer readable medium article of manufacture with instructions for the purpose of preparing a relational database for data-mining as a flat database, the medium comprising instructions that when executed:

identify a data model;

generate a data hierarchy tree;

collect multiple events in child records associated with a parent record;

characterize the nature of multiple events in the child record;

extract features from the child records, where feature extraction depends on the nature of the multiple events in the child records;

append extracted features to the parent record; and

repeat the method for all child records.

29. A computer readable medium article of manufacture with instructions for the purpose of transforming a relational

database to a flat database, the medium comprising instructions that when executed:

provide a relational database having a first table and a second table, each table having a plurality of records, each record having a plurality of fields, wherein a linked field in a selection record in the first table contains data corresponding to data in a linking field of a plurality of records in the second table;

characterize the data in a summarized field in the second database by computing summarization data, wherein the summarized field in the second database is not the linking field; and

append a summarization field to the first table; and store the summarization data in the summarization field of the selection record in the first table.

30. .The medium according to claim 29 further comprising instructions that when executed repeat the characterizing step and the appending step for all records in the first table.

31. A computer readable medium article of manufacture with instructions for the purpose of applying a data mining technique for a flat database to a relational database, the medium comprising instructions that when executed:

provide a relational database having a parent table, parent-table records, a child table, and child-table records, wherein a plurality of child table records can be linked to a parent table record;

convert the relational database to a flat database by appending to a parent table record at least one field summarizing the values in child table records linked to the parent table; and

apply a flat database data mining technique to the flat database.

32. A computer readable medium article of manufacture with instructions for the purpose of determining the relationships among tables in a database, the medium comprising instructions that when executed:

- 5 identify potential primary key fields;
- determine table hierarchy that identifies tables as parent tables and related child tables;
- explore intra-table data relationships to reduce the size of a data table; and
- 10 explore inter-table data relationships between data in a parent table and data in a child table to that parent.

33. A computer readable medium article of manufacture with instructions for the purpose of identifying potential primary key fields, the medium comprising instructions that

15 when executed:

- identify a redundant field whose name appears in a plurality of tables;
- identify as a parent table a table in which the value of the redundant field is unique for each record. whereby the
- 20 redundant field is a primary key for the parent table;
- select as a parent record a record from the parent table, whereby the value of the redundant field of the parent record is unique in the parent table;
- select as child records all records in tables other than
- 25 the parent table for which the value of the redundant field is the same as the value of the redundant field in the parent record; and
- identify as a child table a table that is not the parent table and that has the redundant field.

- 30 34. A memory for storing data for analysis by a data mining application, the memory comprising:

a data structure stored in said memory comprising a flat database table;

a primary record in the database table reflecting one instance of a set of fields of data, the record being
5 associated with a plurality of secondary records in a linked database table;

a raw data field in the database table containing raw data stored in the table; and

a transformed data field in the database table containing
10 transformed data, the transformed data field in the primary record representing the plurality of secondary records associated with the primary record.

35. The memory according to claim 34 wherein the transformed data field is a statistic summarizing the values
15 of the plurality of records associated with the primary record.

36. The memory according to claim 34 wherein the transformed data field is a computed transformation of the values of the plurality of records associated with the primary
20 record.